LHC Computing Grid Project - LCG

David Foster – LCG Chief Technology Officer

Information Technology Division

CERN – European Organization for Nuclear Research

Geneva, Switzerland

david.foster@cern.ch

The Large Hadron Collider
Requirements for data storage and analysis

4 large detectors

Storage –

Raw recording rate 0.1 – 1 GByte/sec

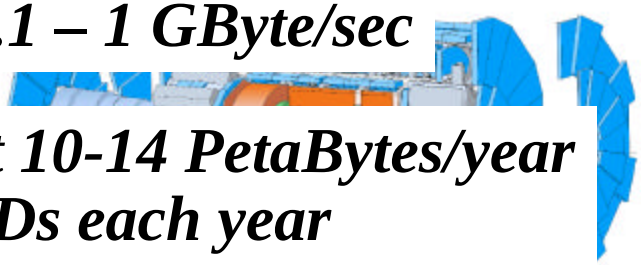
Accumulating data at 10-14 PetaBytes/year
~ 20 million CDs each year

10 PetaBytes of disk

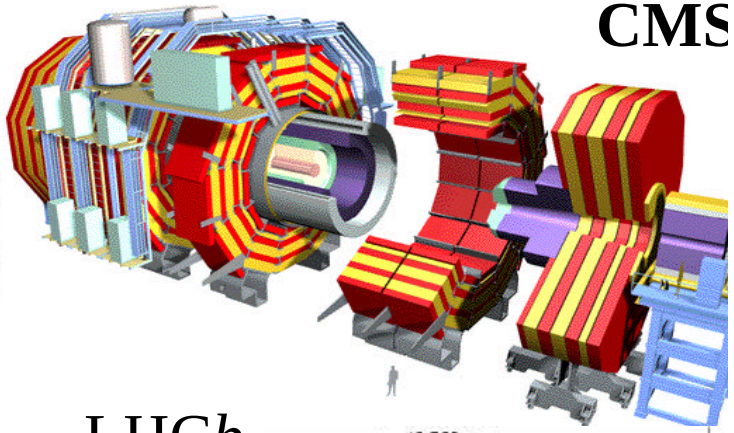
Processing –

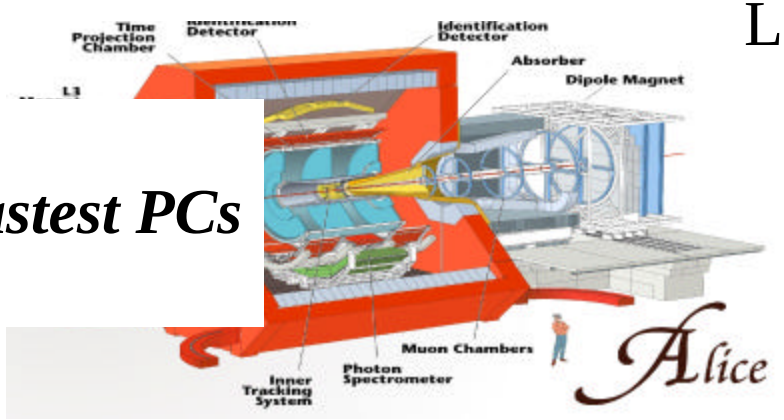
150,000 of today's fastest PCs

TLAS

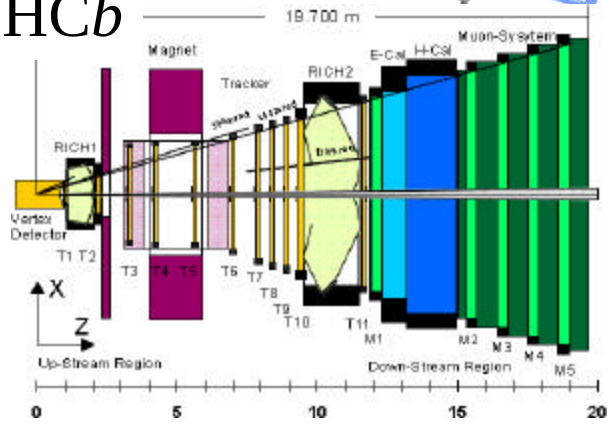


CMS





LHC*b*

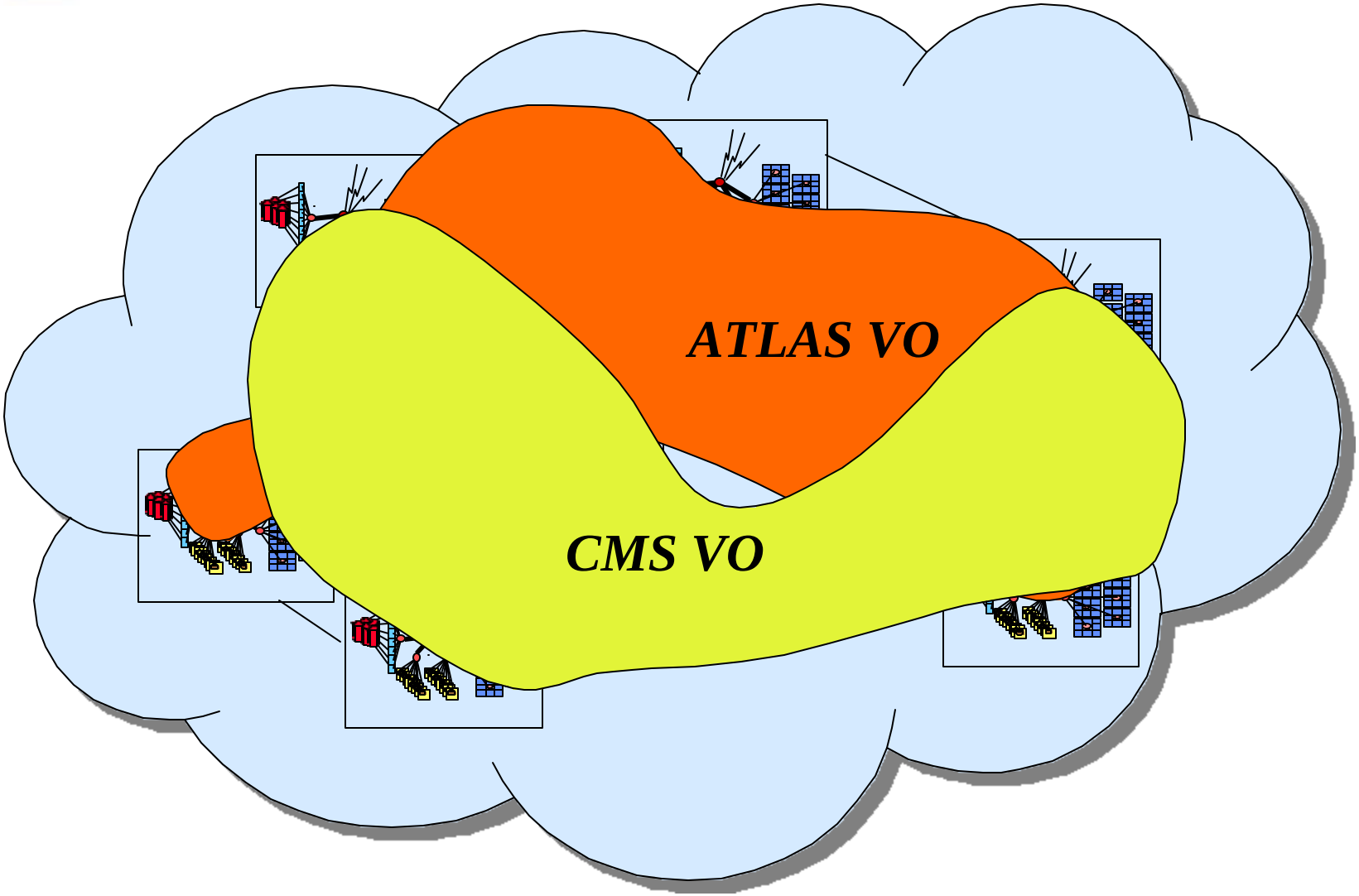


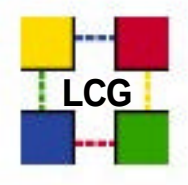
CERN

CERN's Users and Collaborating Institutes

Europe: 267 institutes, 4603 users
Elsewhere: 208 institutes, 1632 users

Grid →The virtual LHC Computing Centre





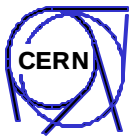
Virtual Computing Centre

The user ---

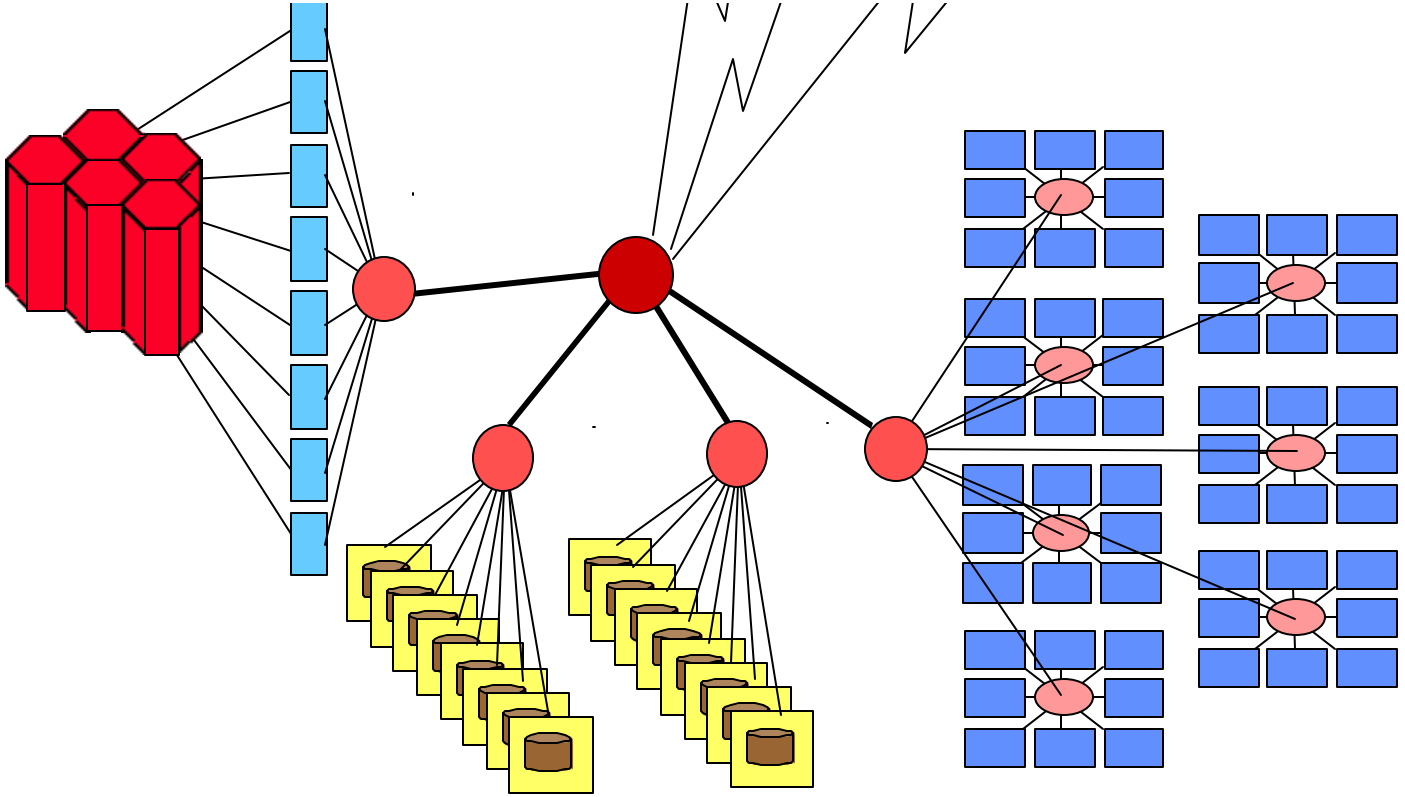
sees the image of a single cluster

does not need to know - where the data is
- where the processing capacity is
- how things are interconnected
- the details of the different hardware

and is not concerned by the local policies of the equipment owners and managers

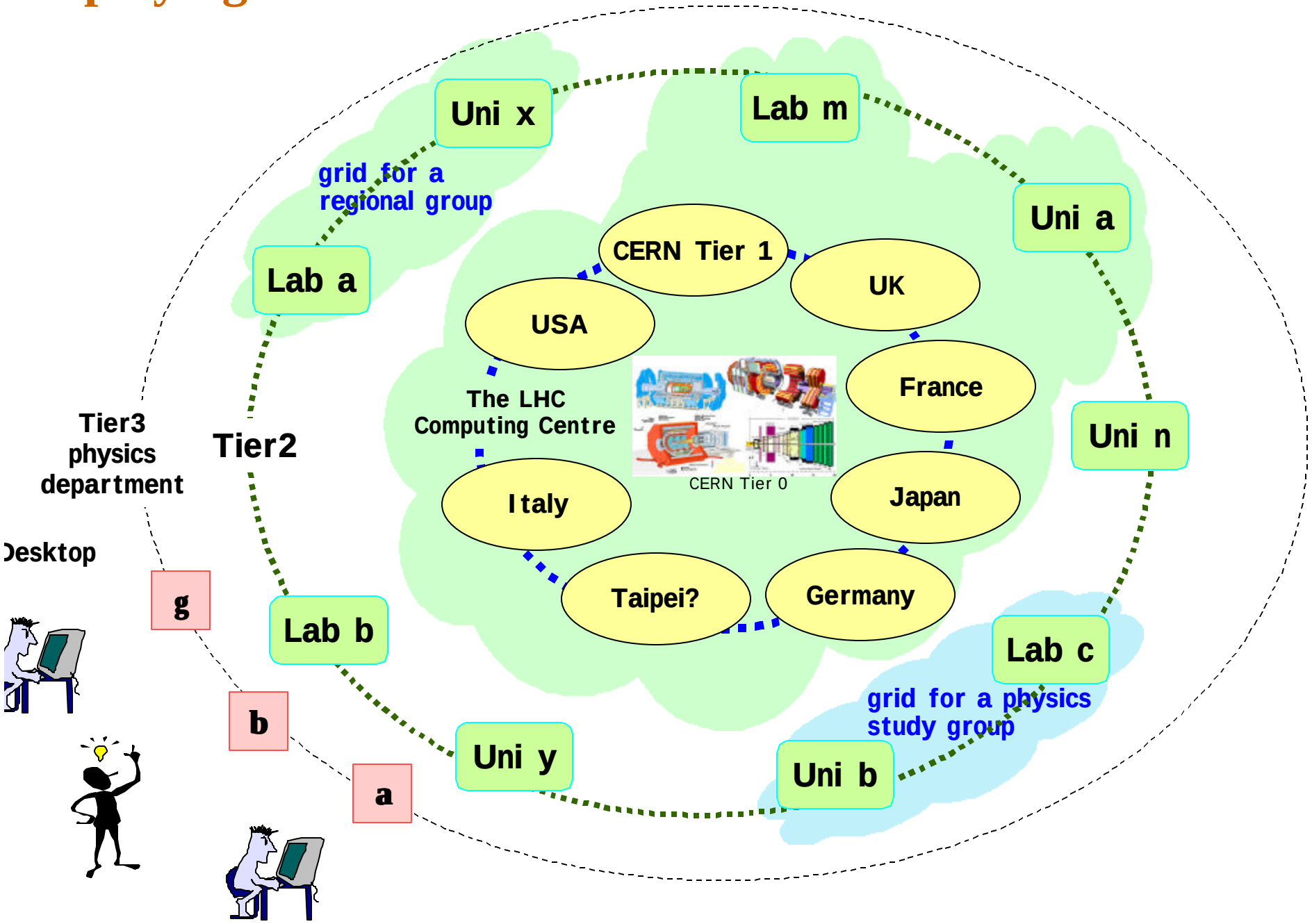


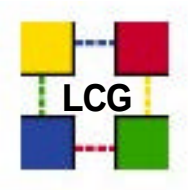
Each Site has physical, logical and policy constraints.
Sites may/typically have well defined small number of connections to the outside world.



Sites Serve Multiple Application Groups - LHC and Other

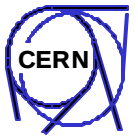
Deploying the LHC Grid

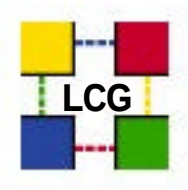




LCG Activities

- Applications
  - Understanding the application issues including common requirements, development strategies, physics data object handling etc.
- Fabrics
  - Working towards the management of large fabrics. What are the fabric strategies including disk and CPU services.
  - Working as part of the CERN Openlab with Industry.
- Grid Technologies
  - Working towards interoperable solutions between Europe and the USA
- Deployment
  - Organising the relationships with the Tier-n centers including testing, deployment and management methodologies.





Deploying the LCG Service

First Pilot - Target June 03 –

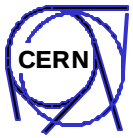
deploy a **Global Grid Service**

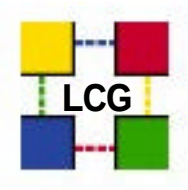
- data replication, migration
- sustained 24 X 7 service
- including sites from three continents
- several times the capacity of the CERN facility
- and as easy to use

learn how to –

-- manage large distributed data collections

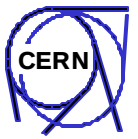
-- operate, maintain and support a grid



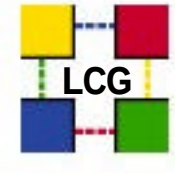


Issues Being Addressed

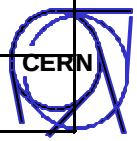
- Interoperability of middleware initiatives
  - GLUE activities (Thanks to Ruth Pordes for the material)
- Technical Grid Deployment
  - Certification and Packaging
- Regional Center Agreements
  - Grid Deployment Board First Meeting
    - 14 Countries + Experiments present
  - Agree Policies and Procedures
  - Operations management, change procedures
  - Resource allocation, accounting
  - Security Issues
  - Problem resolution and user support procedures

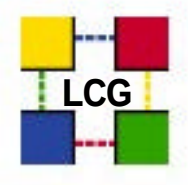


The Glue project - the people involved "day to day"



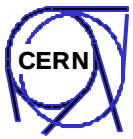
Sergio Andreozzi	DataTAG	Schema, Glue testbed	Carl Kesselman	iVDGL, Globus	Schema
Olof Baring	EDG WP4	Schema, Information providers	Peter Kunszt	EDG WP2	Schema, Data Movement and Replication
Rick Cavanaugh	GriPhyN, iVDGL	Applications	Doug Olson	PPDG	Authentication, Authorization
Roberto Cecchini	EDG, DataTAG	Authentication, Authorization	Ruth Pordes	PPDG, iVDGL	Testbeds, Applications
Vincenzo Ciaschini	DataTAG	Glue testbed, job submission	David Rebatto	DataTAG	Applications
Ben Clifford	iVDGL, Globus	MDS development	Alain Roy	iVDGL, Condor	Virtual Data Toolkit packaging, support.
Ewa Deelman	iVDGL, Globus	Schema, VO Operations	Dane Skow	PPDG	Authentication, Authorization
Luca Dell'Agnello	DataTAG	Authentication, Authorization	Scott Gose	iVDGL, Globus	Testbed operations, Glue validation tests
Alan DeSmet	PPDG, Condor	Applications	Massimo Sgaravatto	EDG WP1	Schema, Job Scheduling
Flavia Donno	EDG, DataTAG, LCG	Applications, Job Submission, data Movement	Jenny Schopf	PPDG, iVDGL, Globus	Schema, Monitoring
Sergio Fantinel	DataTAG	Applications	Arie Shoshani	PPDG, LBNL	Storage Interface (SRM)
Enrico Ferro	DataTAG	Distribution, Applications	Fabio Spataro	DataTAG	Authentication, Authorization
Rob Gardner	iVDGL	Applications, Testbed	Regina Tam	EDG WP5	Schema
Jerry Gieraltowski	PPDG	Applications	Brian Tierney	PPDG, LBNL	Schema, Monitoring
John Gordon	EDG WP5	Storage Schema and Services	Luca Vaccarossa	DataTAG	Applications
David Groep	EDG	Authorization	Cristina Vistoli	DataTAG	Schema, Coordination
Leigh Grunhoefer	iVDGL	Authentication, Testbed	Saul Youssef	iVDGL	Software Distribution, Applications

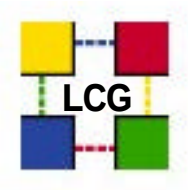




Glue: Components of Interoperability

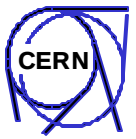
- Authentication
- Authorization
- Job Submission
- Resource Discovery
- Job Description and User Interface
- Storage Management
- Data Movement, Access, and Handling
- Software Distribution and Package Management
- Validation of Interoperability

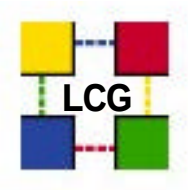




Glue: Validation of Interoperability

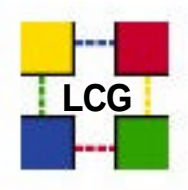
- **In Place today: Demonstrations showing interoperability at one time and location:**
  - Component level tests - AA, GRAM, GridFTP, GDMP, MDS
  - "two node" application test - MOP, ATLAS
- **In progress for tomorrow: Multi-site application demonstrations with automation in installation and configuration in progress for November conferences**
  - ATLAS simulation
  - CMS MOP and Virtual Data
- **Needed for next year: Ongoing system demonstrations. Formal Validation that needs to be ideally done for each new release of software and system change**
  - EDG test scripts
  - VDT test harness and scripts
- Intergrid Demonstrations for IST2002 and SC2002





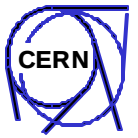
LCG Direction

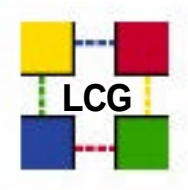
- LHC data analysis has enormous requirements for storage and computation
- Grid technology offers a solution for LHC - to unite the facilities available in different countries in a virtual computing facility
- The technology is immature – but we need reliable solutions that can be operated round the clock, round the world
- The next three years work –
  - set up a pilot service – and use it to do physics
  - encourage the technology suppliers to work on the quality as well as the functionality of their software
  - learn how to operate a global grid



What is needed ?

- To be able to get new sites on the grid
  - High quality software package, best practices.
- To be able to deal with operational problems
  - Failures, Security Problems, Problem determination, Evolution
- To be able to provide high quality services
  - Problem tracking/resolution, user support
- To be able to effectively leverage the resources available – and prove it !
  - Job matching, accounting, auditing
- Manage the evolution
  - Technology, Policies, Standards etc





Where a GGF Production Grid Activity Could Help

- Help Define Middleware Service Requirements
  - Reliability, Robustness,
  - Instrumentation and Control
  - Error Reporting and Handling
  - Facilities for system administrators and operations tools
- Recommend Deployment Strategies
  - Preparation
  - Packaging and Delivery
  - Management and Updates
  - Interoperability Issues
- Kickstart New Sites
  - Collected Best Practices
    - Maintaining Security
    - Configuration Management
    - Support policies and procedures
    - Etc ……